IBM APPLIED DATA SCIENCE CAPSTONE

Opening a New Multiplex Business
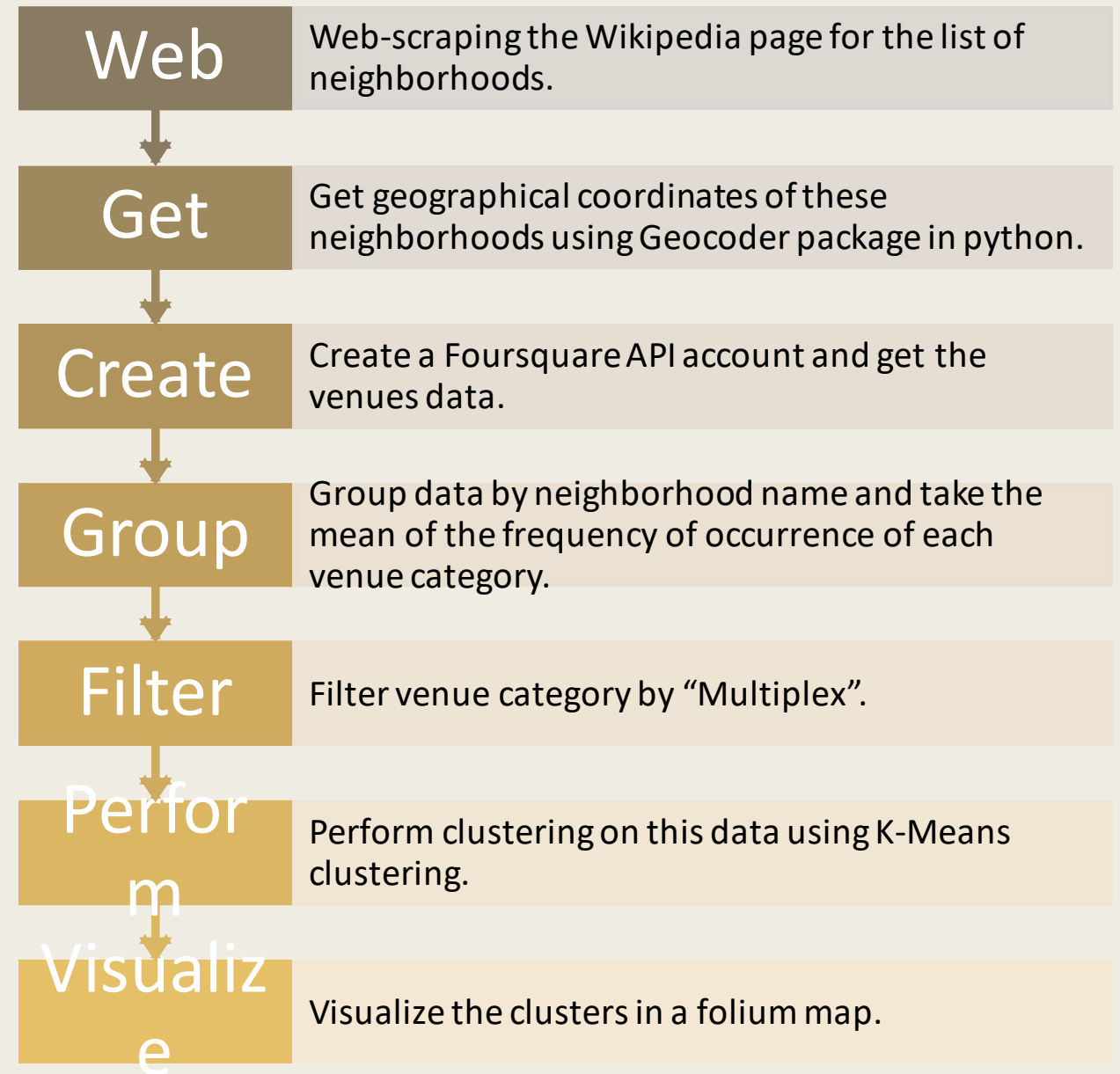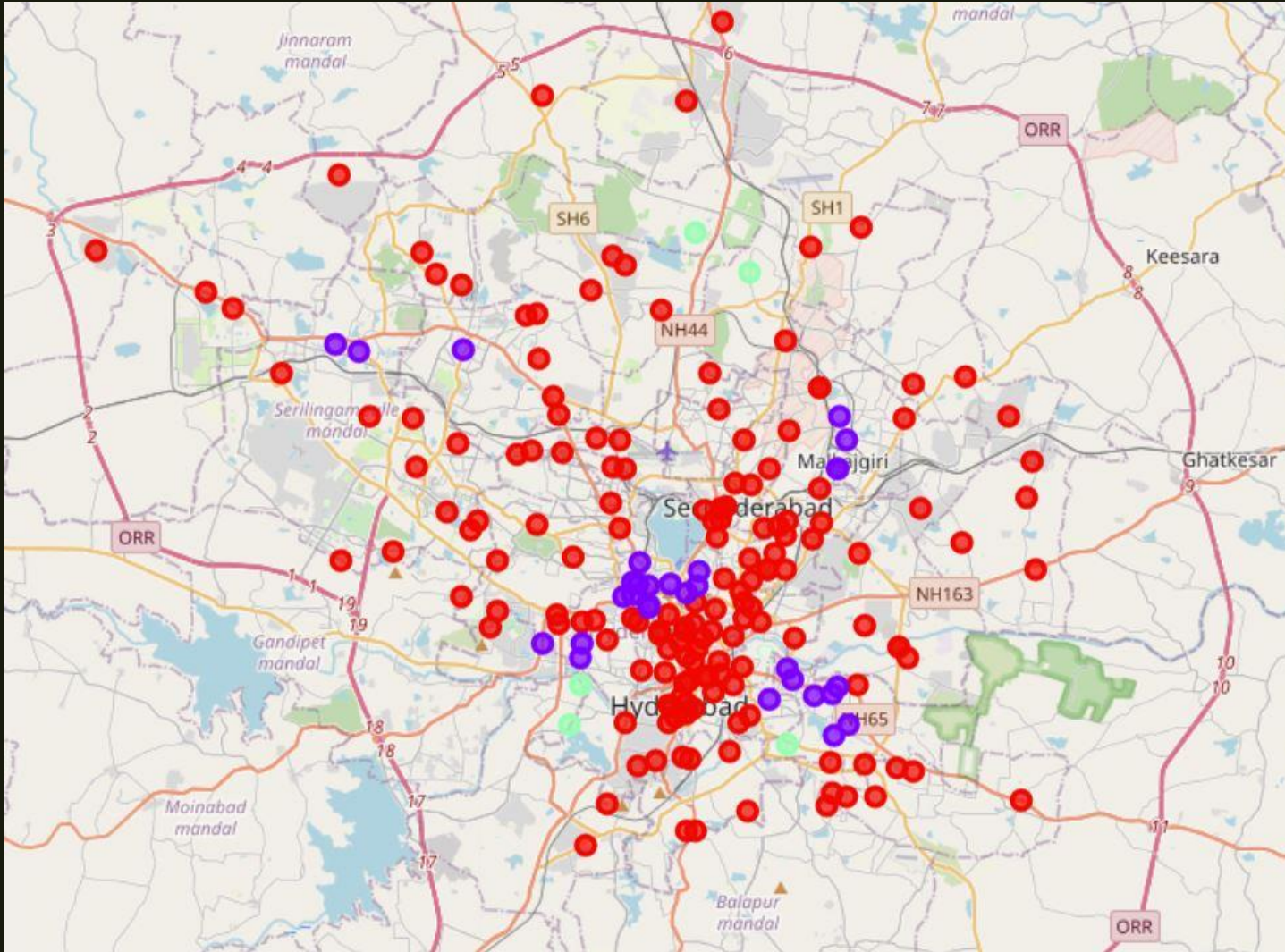
By,

Phalgun Reddy Bobbili

Business Problem

- One of most important factors driving the success of a business is the location.

- The objective of this project is finding the most suitable neighborhoods to start a new multiplex business in Hyderabad, India.

- I chose Hyderabad as the focus because the love for cinemas in India, and specifically in southern India where Hyderabad is located, is very high. People love going to the cinemas to enjoy the big screen experience.

Data

- The data required for this project is as follows:
  - List of neighborhoods in the city of Hyderabad
  - Geographical coordinates of these neighborhoods
  - Venue data of various venues in the city
- The sources of data are:
  - List of neighborhoods

  https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Hyderabad,_India
  - Venues data

  https://developer.foursquare.com/docs/

Methodology

Web	Web-scraping the Wikipedia page for the list of neighborhoods.
Get	Get geographical coordinates of these neighborhoods using Geocoder package in python.
Create	Create a Foursquare API account and get the venues data.
Group	Group data by neighborhood name and take the mean of the frequency of occurrence of each venue category.
Filter	Filter venue category by "Multiplex".
Perform	Perform clustering on this data using K-Means clustering.
Visualize	Visualize the clusters in a folium map.

Results

- Cluster-0 in red:- Neighborhoods with moderate number of multiplexes.
- Cluster-1 in purple:- Neighborhoods with few or no multiplexes.
- Cluster-2 in mint green:- Neighborhoods with high concentration of multiplexes.

Conclusion

The neighborhoods in cluster-1 are the most preferred locations to open a new multiplex business.

The findings of this project alone may not be enough to make decisions. We need to pursue further studies of the demographics, real estate markets, etc. to be able to make a more calculated choice of neighborhood.

THANK YOU

Q and A